BISS Biodiversity Information Science and Standards

Biodiversity Information Science and Standards 2: e26300
doi: 10.3897/biss.2.26300

OPEN ACCESS

Conference Abstract

Putting your Finger upon the Simplest Data

Arturo H. Ariño ‡

‡ University of Navarra, Department of Environmental Biology, 31080 Pamplona, Spain

Corresponding author: Arturo H. Ariño (artarip@unav.es)

Received: 30 Apr 2018 | Published: 15 Jun 2018

Citation: Ariño A (2018) Putting your Finger upon the Simplest Data. Biodiversity Information Science and Standards 2: e26300. https://doi.org/10.3897/biss.2.26300

Abstract

Over the past decades, digitization endeavors across many institutions holding natural history collections (NHCs) have multiplied with three broad aims: first, to facilitate collection management by moving existing analog catalogues into digital form; second, to efficiently document and inventory specimens in collections, including imaging them as taxonomical surrogates; and third, to enable discovery of, and access to, the resulting collection data.

NHCs contain a unique wealth of potential knowledge in the form of primary biodiversity data records (PBR): at its most basic level, the “what, where and when” of occurrences of the specimens in the collections. But as T.S. Eliot famously said, “knowledge is invariably a matter of degree”. For such data to be transformed into digitally accessible knowledge (DAK) that is conducive to an understanding about how the natural world works, release of digitized data (the “this we know”) is necessary.

At least two billion specimens are estimated to exist in NHCs already, but only a small fraction can be considered properly DAK: most have either not been digitized yet, or not released through a discovery facility. Digitizing is relatively costly as it often entails manually processing each specimen unit (e.g. a herbarium sheet, a pinned insect, or a vial full of invertebrates). How long could it take us to transform all NHCs into DAK? Can we keep up with the natural growth in collections?

The Global Biodiversity Information Facility (GBIF) has become the de facto main index of PBR, both originated in NHCs or as field observations. Digitized NHC that are standards-

© Ariño A. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

compliant and can be connected to, or harvested by, GBIF, effectively become DAK. I have examined GBIF growth data looking for a pattern of DAK generation.

I found that the rate of NHC-based PBR accrual is remarkably constant: the total DAK shows a strongly linear growth, as opposed to the exponential growth exhibited by cumulative observation data. Projecting the trend to the estimated holdings shoots the completion many decades ahead. In addition, digitized data appear to be taxonomically biased.

Digitization efforts must therefore step up qualitatively in order to enable processing the backlog, let alone newly-acquired accessions, within one generation. Among several possible solutions, emerging, industrial-scale mass-digitization techniques may help harnessing this otherwise daunting task—but there's also a risk that DAK becomes even more uneven across taxon groups because of the narrow application specificity of such techniques, thus potentially biasing our knowledge of nature.

Keywords

Natural History Collections (NHC), primary biodiversity data records (PBR), digitally accessible knowledge (DAK), trends, digitization, bias

Presenting author

Arturo H. Ariño

Presented at

SPNHC 2018